DOCUMENT. RESUME

ED 418 725 IR 057 026

AUTHOR	Soete, George J.
TITLE	Transforming Libraries: Issues and Innovations in Preserving Digital Information. SPEC Kit 228 and SPEC Flyer 228. Issue Number 5.
INSTITUTION	Association of Research Libraries, Washington, DC.
ISSN	ISSN-0160-3582
PUB DATE	1997-12-00
NOTE	43p.
AVAILABLE FROM	Association of Research Libraries, Office of Management Services, 21 Dupont Circle, Washington, DC 20036 ($28).
PUB TYPE	Reports - Research (143) -- Tests/Questionnaires (160)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Access to Information; *Archives; Information Management; Information Policy; Information Services; *Information Storage; Information Technology; *Library Technical Processes; Models; *Preservation
IDENTIFIERS	Association of Research Libraries; *Digital Data; Stakeholders

ABSTRACT

The problem of preserving digital information and the strategies that are and might be employed to address it are the focus of this fifth issue of "Transforming Libraries." Twenty-one individuals involved at the technical or policy level in developing strategies for preserving digital information were interviewed. There is consensus on a number of points, including: the problem is growing faster than solutions are being worked out; often confusing is the difference between digitizing for preservation and digitizing for access, and preserving digital-only information; and more analysis of the problem, including further assessment of the technical issues, costs, options, and risks needs to be undertaken in a collaborative manner. Ten stakeholder groups who will play a role in designing and testing possible models are: researchers or originators of information; publishers; repositories; libraries and museums; the national libraries and archives who operate legal deposit programs; the cooperative service agencies; private storage providers; the scholarly associations; indexing and abstracting services; segmented market providers; and certified digital archives as proposed in a study by the Commission on Preservation and Access and the Research Libraries Group. The "Reports from the Field" section of this document features some of the varying institutional approaches to preserving digital information. (AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

IR057026

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C.A. Mandel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Preserving Digital Information

Transforming Libraries

Issues and Innovations in

SPEC KIT 228

5

<http://www.arl.org/transform/>

Written by
George J. Soete,
Transforming Libraries Editor

Editorial Advisor
Mary Case,
Director, ARL Office of Scholarly Communication

BEST COPY AVAILABLE

Systems and Procedures Exchange Center
SPEC Kit 228

ISSN# 0160 3582

SPEC Editor Patricia Brennan
Association of Research Libraries

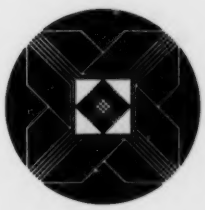
Graphic Designer Kevin Osborn,
Research & Design, Ltd., Arlington, Virginia

Transforming Libraries price per issue: $28

Copyright © December 1997

This volume is copyrighted by the Association of Research Libraries. ARL grants permission to reproduce and distribute copies of this work for nonprofit educational or library purposes, provided that copies are distributed at or below cost, and that the author, source, and copyright notice are included on each copy. This permission is in addition to rights of reproduction granted under Sections 107, 108, and other provisions of the U.S. Copyright Act.

∞ *The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Performance of Paper for Printed Library Materials, ANSI Z39.48-1992.*



ASSOCIATION OF RESEARCH LIBRARIES
OFFICE OF MANAGEMENT SERVICES
21 Dupont Circle
Washington, DC 20036
Phone (202) 296-8656
Fax (202) 872-0884
<http://www.arl.org> <pubs@arl.org>

Transforming Libraries *Issues and Innovations in* **Preserving Digital Information**

5

BEST COPY AVAILABLE

Table of Contents

Introduction: A Word from the Editor 5
The Problem 6
The Strategies 7
Key Models 8
Issues Roundtable 10
Next Steps 13

Reports from the Field 16
UNIVERSITY LIBRARIES 17
Agricultural Literature in Good Hands with Cornell/NAL Partnership 17
Harvard Has Entered the Design and Planning Stages of Digital Library Development 20
Michigan Adopts Strategic Combination of Preservation and Access 21
COOPERATIVE MEMBERSHIP ORGANIZATIONS 22
ICPSR Has Long History of Archiving Electronic Information 22
OCLC Focuses on Members' Current Needs in Digital Preservation 23
RLG Experiments, Develops Infrastructure 25
NATIONAL LIBRARIES AND ARCHIVES 27
LC Coordinates Management of Electronic Resources 27
National Library of Canada Focuses on Policy, Acquisition Procedures 28
CER/NARA Mission: To Provide Ready Access to Essential Digital Evidence 29
LEGAL DEPOSIT ARRANGEMENTS 30
PUBLISHERS 32
American Institute of Physics Plans for Perpetual Preservation 32
UMI Dissertations is Recreating its Business for the Digital Age 33
OTHER COORDINATIVE EFFORTS 34
Peter Graham of Rutgers Plans ARL/CNI/CLIR Workshops, Initiates Local Discussions 34
Digital Library Federation Moves to Coordinate National Effort 35

Additional Resources 37

Introduction: A Word from the Editor

WHEN WE CONFRONT the problem of preserving digital information, we confront the very essence of what it will mean to be a library in the 21st century. The problem has all the hallmarks of the sort of challenge that libraries find both monstrously intimidating and compellingly attractive: danger, opportunity, technical complexity, paradox, unknowable answers, and apparent insolubility. It has compelled a number of major stakeholders to extend intensive planning efforts and take decisive actions. It has yet to hit most of us like the ton of bricks that it truly is.

Film maker Terry Sanders effectively captures the flavor and texture of this problem in his recent film, *Into the Future*. In a very early sequence, a truck loaded with hazardous waste is headed toward a dump site. The question is asked: Will our descendants know where we have buried this waste? Will we have archived the labyrinthine computer records of this and other sites responsibly so that hundreds of years from now the hazardous wastes will not pose a lethal threat? The narrator reminds us of our "obligatory one-way monologue with the future." And though libraries themselves may not be held responsible for such life-and-death archiving (or will they?), they will surely be responsible for a major part of the monologue.

How will we transmit research results, official evidence, and our cultural heritage to future generations? In large part, that transmission will occur in the digital streams that we preserve, along with the tools for making sense of them. There are so many clear advantages in digital preservation that everyone simply assumes its wisdom. But, although short-term digital storage is familiar enough—we save personal documents to disks for later use and we maintain huge databases through regular backup procedures—digital preservation hands us the challenge of preserving a document or a database or a dataset or a computer game for fifty, one hundred, or even five hundred years.

The Problem

Until now, we have perhaps underestimated the advantages of a largely print culture. Though paper can burn or be consumed by inherent acids, if reasonably well made and cared for, a print document can last for centuries. So can preservation-quality microfilm. With digital information, there are significant dangers of physical deterioration of storage media. Magnetic media are particularly subject to data loss and errors. Beyond physical deterioration is an even more daunting problem: hardware and software obsolescence. In fifty years, will we have the hardware on which to mount the physical medium? (Probably not.) And supposing we regularly copy data onto new media, will we have the software needed to make sense of the data? (Again, probably not.)

The technical problems are further complicated when resources to be preserved exist in digital form only, especially if they are dynamic—that is, still online and regularly evolving. Most digital preservation projects today actually involve digitization of paper analogs for access and preservation. Though the technical problems can be very similar, digitizing a journal or scanning a photograph and holding onto at least one hard copy analog is much less risky than preserving digital information for which there is no durable analogous safety net. Beyond the technical problems (which most experts believe will be solved eventually) there is a host of financial, legal, and policy problems. Who will pay for preserving digital information? Who will decide what to preserve, on what media, and with what level of metadata management? How will we, as a society, organize ourselves to make sure that everything that needs to be preserved is preserved?

And then there is copyright. Widespread uncertainty about legal requirements for managing intellectual property in digital environments has presented a significant barrier to the preservation of copyrighted digital resources. One of the chief reasons digital archiving has focused on the digitization of older collections is that they are in the public domain. Even national libraries face obstacles when they try to archive and service their own governments' publications.

All of these problems converge in special ways for libraries at a time when few have received any recent increases in resources, much less the additional resources needed to mount a digital preservation program. In fact, the digital imperatives that concern most libraries are much closer at hand: mostly, they focus on access rather than preservation. The problem is usually: How will they subscribe to and provide access to a critical mass of electronic journals?—a budgetary problem, chiefly, rather than a technical one.

How can they use digital technologies to provide special collections finding aids and image collections over the Internet? How can they provide electronic reserves services to distance education users? How can they offer geographic information services? How can they afford the infrastructure needed to do all of this? Surely, many libraries feel, the full challenge of preservation of purely digital information does not fall solely on libraries.

While some are at work on the problem of digital preservation, and while doing something concrete and local about the problem may be beyond the present reach of many libraries, it is a problem that will require thought and support from the broadest possible community. This issue of *Transforming Libraries* seeks to summarize some of the strategies and models currently being explored, and reports as well on several specific efforts coming out of national libraries and archives, university libraries, government agencies, non-profit cooperatives, and professional associations. It is not meant to be a comprehensive survey; rather it is intended to be a starting point or stimulus for further study and work at the local level.

The Strategies

The purely technical strategies for digital preservation that are most often used are refreshing and migration. Most experts agree that refreshing data (copying it periodically onto more stable media) cannot solve the long term problem. Though you might be able to save simple ASCII files this way, for anything more complex you are likely to lose whatever critical functionality was built into the original file, functionality that may be so basic that the data won't be understood without it. There are strong partisans in the field, however, for migration and for a particular sort of migration called emulation. Migration and emulation are familiar enough: when we get a new version of a word processing system, we migrate older files to the new system; we are able to do this because the new system has been designed to emulate the old system so that it can accept the old files. Migration focuses on fitting preserved data that might arrive in a vast array of formats into a smaller number of formats that can still encode the complexity of structure and form of the original. Emulation focuses on designing hardware and software that will emulate older programs and thus reproduce their functionality. The strategy of choice for most current archival operations appears to be regular migration onto magnetic media.

Another strategic approach is for a digital archive to develop standards and policies that stipulate the formats that data will be accepted in and/or restrict acquisitions in other ways to assure the ability to sustain the archival mission. In the analogous "print

culture," such strategies have been embedded in library collection and archival retention policies for years. Thus, an electronic archive might, for example, accept text-based files but not computer games, or published works but not e-mail. A related strategy is to provide advice and training to those who submit data, to assure that they will meet standards and policy requirements. In these ways, the archiving organization promotes the development of standard practice and seeks to limit the problem cases that must be dealt with.

Several players in the area of digital preservation are engaging in strategic speculation—building a technical infrastructure and trying various economic models. To test the infrastructure and to assess costs, they are mounting test collections, such as RLG's Studies in Scarlet and OCLC's Digital Schomburg Collection. Even though these are print-based collections, the projects seek to answer fundamental questions: Will such collections be used? Will there be a viable market for digital preservation services? How will preservation and servicing of collections be paid for—through assessment of members, or per use charges? Carefully monitoring the RLG and OCLC experience in answering these questions will help others organize digital preservation efforts.

The Key Models

As one investigates the myriad ways that preservation of archival information occurs, several models become apparent. For a while, perhaps indefinitely—and perhaps eventually in a seamless international structure—all of these models will be with us. For now, they exist in a colorful array of possibilities. The following is not meant to be an exhaustive list, nor are the models mutually exclusive. One key point was made by several persons interviewed for this issue: whichever models eventually prevail, libraries and their users will pay for digital preservation, either directly or indirectly. It is critically important, therefore, for libraries to get involved, pay attention to technical and policy developments, and support those that are likely to benefit libraries' users most.

1. **Originators**

   Much current "archiving" is done by individual researchers, research and development units, etc., and many researchers wonder why they would have to depend on others to archive data. As long as research results are merely shared among a handful of colleagues, this approach has a proven track record. But, clearly, it is not fail-safe and it can leave much to be desired if we expect universal accessibility. What happens when the research project folds, or when the researcher retires or dies?

2. **PUBLISHERS**

   This model, already put in place by a number of publishers, assumes that publishers will provide perpetual preservation and access as part of their contract with subscribers. Such archives will also typically be available on a per transaction basis to non-subscribers. Archives might also include so-called auxiliary materials: for example, the data behind a research publication. Many librarians express doubts about this model. What will happen if the publisher ceases business? What will happen if maintaining electronic archives turns out to be unprofitable, or even a money-loser?

3. **LIBRARIES, MUSEUMS, AND OTHER CONSERVATOR INSTITUTIONS WITHIN A DISTRIBUTED NETWORK**

   A few libraries are taking steps in this direction already, especially in the storage of digital information that has a tangible analog (e.g., books and journals). The Mann Library at Cornell, featured later in this publication, has taken responsibility for archiving research publications in the field of agriculture. Many others are creating digital storage for their own special collections materials (though how many of these are actually "preservation" efforts in addition to service programs is an open question). Mechanisms for national and international coordination will be needed to make sure that such schemes cover what we need to preserve.

4. **NATIONAL LIBRARIES AND ARCHIVES THROUGH LEGAL DEPOSIT PROGRAMS**

   Depository legislation for electronic information varies widely among nations that have depository arrangements. Though this is an attractive option, there are barriers to overcome, not the least of which is copyright, even for publications originating in the library's country. Can any national library truly commit to acquiring by deposit all electronic publications that should be acquired? Will such programs be subject to the vagaries of budget appropriations?

5. **COOPERATIVE SERVICE AGENCIES**

   OCLC and RLG have already begun to position themselves to provide digital storage services for members if a market develops. As noted later in this publication, they are developing infrastructure, mounting special collections, and making electronic versions of journals available to users. A big question is how economically viable storage of digital-only information will be. Will members be willing to pay for such services through membership fees? Will depositors pay? Will users pay?

6. **"SEGMENTED MARKET" PROVIDERS**

   UMI and ICPSR are current examples of "segmented market" providers, with UMI "responsible" for preserving dissertation

literature and ICPSR for preserving social science research data. A key advantage is that such providers get to know their "product" and operation very well, developing impressive economies of scale and value-added services as they grow and learn. Are there other market segments for which it might make sense—both logically and economically—for other providers to take on?

7. **PRIVATE STORAGE PROVIDERS**
   As with preservation microfilming services, will there develop an industry that provides digital storage for a fee? Again, how dependable might such services be if they are subject to the whims of the marketplace?
8. **SCHOLARLY ASSOCIATIONS**
   Associations such as the American Institute of Physics are planning to archive their own publications. Might they, within their research areas, become depositories or service agencies for purely digital information, as well?
9. **INDEXING AND ABSTRACTING SERVICES**
   There could be a natural connection between providing links to digitally archived research and archiving the research itself.
10. **CERTIFIED DIGITAL ARCHIVES.**
    The 1995 RLG/CPA Report (see list of resources) recommends that digital archives be certified to assure acceptable levels of quality.

Issues Roundtable

Twenty-one key stakeholders in the effort to preserve digital information were interviewed for this publication. Much of the information they provided about their own activities and operations can be found in the Reports from the Field section. Their more general comments on the issues are summarized here.

In addition to the people listed as contacts in the Reports section, the contributions of Margaret Hedstrom, Clifford Lynch, Deanna Marcum, and Jeff Rothenberg are gratefully acknowledged.

The Problem: the Current Situation

There are fears. Most interviewees find that the problem is growing faster than we are currently working to solve it. One U.S. librarian felt that the U.K. and Australia seemed to be farther along than the U.S. in dealing with the preservation of digital information on a national policy basis. Another characterized current efforts as still very much in "working group" mode. Yet others noted that, in spite of the excellent work represented in such efforts as the RLG/CPA report, little substantive progress on matters such as the certification of digital archiving services has been made. At

the individual library level, one person commented that those in the current avant-garde are probably those who have merely done a basic inventory of digital resources on their campuses. What is being lost while we get organized? What are we doing about the significant "gray literature" to be found in abundance on the Internet—a culture in which preservation of the past is not well established as a value? What are we doing about the corpus of material that represents very complex technical problems: computer games, for example? And what about the immense amount of material that is routinely and casually captured in temporary media—TV lectures coming out of universities, meetings of Boards of Regents, and convocations of critical national groups?

Several interviewees mentioned the unfortunate confusion among digitizing for preservation, digitizing for access, and preserving digital-only information. Most digital projects in libraries, as noted above, are in fact of the first two sorts, and many focus on materials that are in the public domain. Little attention, outside of a few notable instances, is being paid to the preservation of information that will probably never exist in any but digital form—sets of research data, for example. Many libraries have, according to one person, given a great deal of thought and effort to building digital collections, but little thought to preserving these digitally reformatted resources. In addition, another felt that the uncertainty over copyright is often an impediment to preserving important digital-only materials.

One assessment pointed out some critical gaps: First, there exists a lack of expertise—in part a lack of technical skills, but also a relatively weak understanding of the options, costs, and risks involved. Moreover, financial and infrastructure resources are thin for most libraries. Finally, there is an insufficient understanding of how very different the digital world is, including the fundamentally different behaviors and approaches it calls for. These lacks led one person to cite the need for an in-depth survey of the current state of digital preservation.

While many alluded to the problems, a few interviewees suggested that librarians were, in general, unduly intimidated by the problems of digital preservation, and that we have to undergo a philosophical shift that will enable us to see the problem as soluble and then to set about solving it.

Choices

Making conscious choices, through policy development and standard setting, was a constant theme in the interviews. Several people urged choices of durable technologies. For one person, relational databases and magnetic media were preferable because of their durability. Another found PDF files suitable for servicing electronic reserves collections, but not for preservation purposes. For another, the impulse to shape the standards that are used to create digital information in

the first place was analogous to the benevolent pressure put on publishers to use acid-free paper, with a major difference—there are many more creators of digital materials than of paper publications, and thus it is a much greater challenge. Yet another warned that choosing technical standards prematurely would be a disaster; what is needed instead are standards of functionality and performance.

The most important choices, however, happen at the materials selection level. No one argued for the "vacuum cleaner" approach of, for example, trying to capture what is on the entire Internet. One person suggested that librarians and archivists would come closer together in their approaches to selecting digital information for preservation, with librarians in particular becoming more selective about what to preserve.

Several people commented on the negative effect that the attractiveness of the digital world might have, especially when pressure from parent institutions is applied on libraries to "go digital." This will be especially troublesome if libraries choose to digitize for access instead of acquiring, conserving, and preserving in less dazzling but currently more effective ways—in microform, for example.

Models

Skepticism characterized some of the comments about models. One person doubted that any government strategy would be able to solve the problem, with the possible exception of a government taking responsibility to preserve its own digital publications. There was hope, however, that government agencies would provide funding for others to carry out digital preservation. Likewise, there was some question whether those in the private sector who promised "perpetual care" fully understand what it means and really intend to follow through, even if they go out of business. One person felt, however, that such providers could be negotiated with and suitable arrangements would eventually be worked out.

One interviewee suggested looking to analogous efforts. What is being done, for example, about radio and television broadcasts? Are some broadcasts being preserved in well-known projects? Is coverage selective or comprehensive?

Some libraries are considering models in which they might have new critical roles. For example, might a library become a vendor of digital preservation services? Might libraries sell public domain images to others?

One person found no clear economies of scale: centralized operations could not, at this point, be depended on to be cheaper than local efforts. Another felt that we might be investing prematurely in the money-saving aspects of digital storage, especially when strategies such as computer output microform for text might in fact be much cheaper and more durable.

Another person related the economics of digital preservation to the use of the item or data. If something will be used, its use will probably be paid for, and so will its preservation.

Collaboration

Opinions here were fairly decisive. Only the very largest and wealthiest institutions might be able to operate solo when it comes to digital preservation. In the main, single libraries will not be major actors: they will need to work consortially. One person suggested that all of the successful projects have been collaborative. This is certainly the case with the most notable efforts to date (see the Cornell/NAL and ICPSR reports below).

At the same time, suggested one interviewee, we are moving toward a distributed system, and the independence and local creativity of individual participants will continue to be critical. Another suggested that a current problem with many coordinative efforts is that the same institutions have been talking to each other: the dialogue has to widen significantly.

A special issue was mentioned by several people: the need to bring more computer scientists into the effort. One characterized computer science as having been successful in part because of its "luxurious ignorance of the past." The field has, by and large, chosen not to deal with the problem of archiving, mainly because it is simply not as interesting as developing the wonders of the future. Enough computer scientists will eventually become interested, however, to make a difference.

Next Steps

What can ARL and its member libraries do about preserving digital information?

During the October 1997 membership meeting of the Association of Research Libraries, nearly a full day was devoted to a program on the preservation of digital information. Members heard and reacted to six presentations, and, at the end of the presentations, they were given an opportunity to suggest what ARL and individual libraries might do to address the issues inherent in preserving digital information. Moreover, one of the key questions asked of interviewees for this issue of *Transforming Libraries* was, "What can libraries do to advance the cause of digital preservation?" The ideas that emerged in both the interview process and the membership meeting discussion are reflected below.

Potential Next Steps for ARL

Many participants agreed that ARL should take a leadership role in the preservation of digital information in research libraries, providing

support for initiatives underway at individual libraries. Such support might take the form of keeping members informed about projects, perhaps through a web site; fostering a national and international legal framework for intellectual property management; assisting members in breaking the problem into manageable parts; and developing operational projects that solve one problem at a time.

One recommendation for ARL was to create a task force specifically to address the recommendations in the CPA/RLG report that systems be established to certify digital preservation operations and create fail-safe preservation and storage. Such a task force might also include publishers and other providers who are actively developing relevant products and services. At a more global level, it is important for ARL to help develop a national strategy for addressing digital preservation issues by working with CLIR, LC, NARA and other appropriate organizations. Co-sponsorship with organizations such as CLIR, ICPSR, or CNI of a conference including publishers, consortial partners, libraries, and utilities was also suggested.

One specific action suggested was the development of a taxonomy of digital preservation problems, along with specific strategies for addressing them. Another person expressed this notion as a problem matrix and assignment of responsibilities for addressing each problem through assessment of the technical difficulties; exploration of models—how libraries and vendors are solving the problems; and development of action plans. Such an approach would provide assurances that specific technical problems, such as the preservation of web links in otherwise static documents, would be addressed.

Though one member urged that paper collections not be ignored in the process of focusing on digital collections, another noted that paper collections were already receiving most of our attention and that we needed to be sure that the problems represented in digital preservation were addressed.

Potential Next Steps for Individual Libraries

One member encouraged a process of building on a strong history of developing partnerships. Another elaborated on the need to acknowledge the huge scope of the problem and to work more closely with various campus constituencies to develop a framework for addressing this problem, which will have a significant impact for years to come on library management.

The partnership theme was also echoed by many of the experts interviewed for this issue. One suggested that libraries contact their consortia to find out if there were projects that could be joined. Another suggested tapping the expertise in the private sector (one library, for example, works closely with vendors of scanning equipment). Another suggested bringing campus stakeholders into

local discussions of digital preservation, with special focus on campus computer scientists.

Developing strategies for staying informed was another common theme: libraries need to keep track of how various models for preserving digital information are evolving, what colleague institutions are doing, and what standards are being developed. They need also to become active partners in the information gathering process by asking appropriate national organizations to provide guidance—for example, on the tradeoffs between mounting your own digital preservation effort and using a provider.

Finally, there was a set of "mindset" suggestions: to step back, consider the organizational mission and how it might best be served through digital preservation programs; to recall and understand what is known already in the organization about digital preservation—the experiences everyone has had, for example, with migrating digital data from one system to another; to consider analogous experiences with preserving the print heritage—what are the lessons to be learned?; and to work against instinct and intuition by assuming that digital preservation, while difficult, is doable.

Reports from the Field

THE ORIGINAL FOCUS of this publication was on the preservation of digital information that has never existed and probably never will exist in any but digital form. As the interviews progressed, it became clear that there was much less activity in the "purely digital" arena than had been anticipated, that most current digital preservation is in fact digital capture of tangible pieces such as books, journals, photographs, newspapers, and pamphlets—all valuable activities, but not "purely digital."

The decision was made to retain the reports that focus on digitization of tangible analogs for three reasons. First, the technical problems are almost identical; the key difference is that with purely digital information there is no "safety net"—no analogous tangible version that you can recapture if you lose the digital version. Second, as a whole, all of the activities described represent what is currently going on in the digital preservation environment. Third, in virtually every case, the organizations featured are moving toward purely digital preservation if they are not already there, thus, they are using their experience in digitization projects to test a number of factors, including infrastructure and economic models.

As you read through the Reports, therefore, be mindful that while they all represent innovative approaches to digital problem solving, they also represent a mix of activities and experiences: some organizations have a substantial track record in digital-only archiving (for example, ICPSR, Cornell/NAL, and NARA), while others are in intensive planning mode. As one interviewee commented, those in the avant-garde in digital preservation might be those who have merely done an inventory of what needs to be preserved in their environment.

Information professionals in 14 institutions were interviewed for this report. Accounts of what they are planning and doing follow. The reports have been divided into the following categories: University Libraries, Cooperative Membership Organizations, National Libraries and Archives, Legal Deposit Arrangements, Publishers, and Other Coordinative Efforts.

University Libraries

Agricultural Literature in Good Hands with Cornell/NAL Partnership

The Mann Library at Cornell University and the National Agricultural Library (NAL) have taken a leadership role in coordinating activities of the "National Preservation Program for Agricultural Literature" (NPPAL), a discipline-based preservation plan adopted by the United States Agricultural Information Network (USAIN) in 1993. Both institutions have provided staff and funds to assist the program's movement forward. USAIN has also provided funding and a framework that allows USAIN member institutions to use collective talents and resources to further support this national preservation effort. A major current effort focuses on the preservation of digital information in the field of agriculture.

The defining characteristics of the NPPAL are that it: 1) is a national cooperative plan; 2) is discipline based, looking beyond the preservation needs of individual collections to the needs of the discipline as a whole; 3) involves scholars in setting preservation priorities; and 4) has a national library (NAL) that serves a central role as an archive for the literature of agriculture preserved through a series of cooperative projects.

A chief recommendation of the NPPAL is the implementation of a series of cooperative projects based on a conceptual framework that divides the literature of agriculture into a number of logical component parts and assigns responsibility for each. An important offshoot of this program is a project to preserve the core agricultural literature, both books and journals, published between 1850 and 1950. Working within the discipline and genre framework, the project has consulted some 600 agricultural scholars worldwide to discover the publication base which must be preserved. The grant-funded project creates three copies--digital, microfilm, and facsimile—for each item. The microfilm and facsimile copies are the preservation copies. Currently, Mann Library, the library in the Cornell University Library System charged with responsibility for agriculture and life sciences, with funding support from NAL, is securing the necessary rights to mount these materials on the Web, where they will be freely accessible to everyone. At this point, 1,200 core books have been put into the system, and Mann Library is just beginning to work on journals.

Work has begun to supplement the NPPAL by moving into digital publication preservation. NAL, with the Economic Research Service of the USDA (ERS), GPO, Cornell University, and the Farm

Foundation recently took the first step in a cooperative venture by convening a two-day national meeting, "USDA Digital Publications: Creating a Preservation Action Plan," held March 3-4, 1997 in Washington, D.C. Cornell's Mann Library organized the meeting, which served as a call to action to develop a preservation strategy for United States Department of Agriculture digital publications.

According to NAL Director Pam André, "the Library has devoted significant attention and resources over the years to preserving and making available the Department's agricultural information in paper and micro formats. However, it is apparent that the preservation and dissemination of digital information is a much more complex undertaking. The advent of a wholly electronic publishing process poses a significant new challenge to the USDA and NAL. NAL and other stakeholders in USDA believe it is essential to develop a comprehensive strategy for preserving USDA digital publications including the allocation of required resources. This strategy will identify what needs to be done in the near term as well as over an extended period of time, the principal USDA agencies affected, and the resources that will be required."

During the March meeting, representatives of groups that have shared interests and responsibility for digital publication preservation began to identify the major elements and requirements of such a plan. A subsequent consultant's report, *Action Plan for the Preservation of and Permanent Public Access to Digital USDA Publications*, has been issued. The plan calls for an inventory of USDA digital information products, development of criteria and related procedures for life-cycle management of those publications, and identification of technical requirements for implementation. NAL is reviewing the plan and identifying actions to move it forward in the USDA. Following USDA administration approval, NAL will oversee implementation of the plan in cooperation with government and non-government partners.

As part of a land-grant university, the Mann Library has for years had a cooperative agreement with the U. S. Department of Agriculture to mount USDA files on the Mann server and make them available to users within a specified amount of time. The recent movement toward a digital publication preservation capability was in part motivated by the realization that the Mann Library could not continue to store the electronic USDA files that it was making available to the rest of the country, because the USDA is a major producer of digital publications.

This is an important plan for two reasons. First, it clearly moves toward preservation of what former Mann Library Director Jan Olsen called "research on perhaps the most important subject in the world—the very food we eat, our shelter, our clothing, our basic physical well-being." Second, when available as a final plan to

the library community, it will provide a very useful model for local planning.

On another front at Cornell, Anne Kenney, Associate Director, Department of Preservation in the Cornell Library, has taken an interest in preservation cost models and has done her own study of Computer Output Microfilm, which in her view is a viable alternative for many preservation storage purposes, especially given some cost studies suggesting that digital storage might be much more expensive than we had at first imagined. Cornell preservation is looking at all the options, including the possibility that they might outsource digital preservation or even create their own internal outsourcing unit. This imaginative planning grows out of "long and extensive experience with obsolescence of every imaginable kind," according to Kenney.

Contact:

Pamela André
Director
National Agricultural Library
301-504-5248
<pandre@usda.nal.gov>

Evelyn Frangakis
NAL Preservation Officer
National Agricultural Library
301-504-6503
<efrangak@nal.usda.gov>

Maria Pisa
Associate Director for Public Services
National Agricultural Library
301-504-5834
<mpisa@nal.usda.gov>

Anne Kenney
Associate Director, Department of Preservation
Cornell University Library
607-255-6875
<ark3@cornell.edu>

Jan Olsen
Associate University Librarian for Sciences, and
Director, Albert Mann Library [until late 1997]
Cornell University Library
607-255-2285

Sarah Thomas
University Librarian
Cornell University
607-255-3689
<set9@cornell.edu>

Harvard Has Entered the Design and Planning Stages of Digital Library Development

Harvard Library is designing a digital library infrastructure to accommodate a wide range of electronic resources, according to Malloy-Rabinowitz Preservation Librarian Jan Merrill-Oldham. The organization of this effort has led to the appointment of a number of groups that will work collaboratively to achieve specific goals, build a shared knowledge base, avoid duplication of effort, and leverage resources. A key objective in the design process is to build a common framework to ensure an adequate level of consistency in describing, maintaining, and providing access to electronic collections.

Among the teams established are the Harvard College Library Digital Initiatives Working Group (focusing on planning, policy, and program administration), the Visual Resources Task Group (developing shared standards for description, scanning, and electronic delivery of visual materials), and the Digital Library Architecture Group (designing a technical infrastructure to support the organization, delivery, and maintenance of electronic files). The HCL Imaging Services Division and the Harvard University Library Preservation Center are collaborating to evaluate scanning equipment, develop quality control measures, and develop work plans for the establishment of in-house digitizing services.

The current thrust at Harvard is both technical and tactical. Consistent with activities at other institutions, the libraries are intent on building a system that is scalable, flexible, and reliable; and that optimizes the researcher's ability to tap the wealth of the collections—both printed and electronic. Strategic discussions about selection for digitizing and the setting of priorities form the backdrop for technical development.

For Merrill-Oldham and Steve Chapman, Digital Preservation Projects Librarian in the Preservation Center, digitizing promises to be one among several viable preservation options. The emergence of new approaches to protecting and/or securing long-term access to materials both complicate decision making and at the same time promise more satisfactory results. In cases where microfilm falls short of the mark, digitizing will sometimes be an ideal substitute. The greater the number of available options, the better the opportunity to support ongoing teaching and research needs. Merrill-Oldham and Chapman characterize the current effort at Harvard as "serious teamwork."

Contact:
Jan Merrill Oldham
Malloy-Rabinowitz Preservation Librarian
Harvard University Library
617-495-8596
<jmo@harvard.edu>

Stephen Chapman
Digital Preservation Projects Librarian
Harvard University Library
617-495-8596
<stephen_chapman@harvard.edu>

Michigan Adopts Strategic Combination of Preservation and Access

A recent Report from the Field in *Transforming Libraries* #3 (*Electronic Scholarly Publication*) focused on the partnerships that the University of Michigan Library has developed with several publishers in developing their digital library. Further conversations with Wendy Lougee, Assistant Director of the University of Michigan Library, reveal an approach to the preservation of digital resources that is balanced with attention to access.

The approach is a combination of just-in-case and just-in-time planning that is adapted to a variety of needs. As an example, text images are captured and stored in rich TIFF page images, which are converted only as needed to GIF images that users display with a Web browser. Thus, behind what the user sees in the image is the preservation image, which is maintained in a durable and migratable format. The same approach is taken with collections encoded in SGML; only the SGML form is stored, and HTML derivatives are created dynamically, based on user searches.

This method is supported by recent studies suggesting that, like most library collections, only portions of Michigan's electronic collections were actually used over a period of time. As John Price-Wilkin, University of Michigan, recently wrote in *D-Lib* magazine, "As materials are added to our currently relatively small digital libraries, increasingly smaller segments receive the sort of focused attention that we associate with actual use such as reading or printing. The extraordinary boon of the electronic format includes the ability to search across large bodies of material, 'touching' many documents along the way(<http://www.dlib.org/dlib/may97/michigan/05pricewilkin.html>)."

Thus, Michigan is able to present millions of SGML-encoded text and TIFF page images, with "nearly all of it presented in

Web-accessible formats through real-time transformations of the source material," avoiding the waste of creating a library of service derivatives that may or may not be used.

Michigan has also developed an automated routine that performs OCR on TIFF images of text and encodes them in simple SGML to improve searching. The University's very innovative approaches are documented in recent issues of *D-Lib*, which are available on the Web <http://www.dlib.org> and at <http://www.lib.umich.edu/libhome/DLI/>.

Contact:
Wendy Lougee
Assistant Director for Digital Library Initiatives
University of Michigan Library
313-764-8016
<wlougee@umich.edu>

Cooperative Membership Organizations

ICPSR Has Long History of Archiving Electronic Information

The Inter-university Consortium for Political and Social Research (ICPSR), located at the University of Michigan, has been archiving data since 1962. According to Executive Director Richard Rockwell, they have learned that the only way to responsively archive electronic materials is to migrate them. Migration projects take place every three or four years; the last one, which ended in 1996, took a full year. Regular, periodic migration assures the continuing functionality and reliability of archived datasets.

Preservation archiving is a first priority at ICPSR, and an essential focus of the organization is Archival Operation, a three-person unit supervised by Janet Vavra. When data sets are submitted to ICPSR by member institutions, Archival Operation gets them first and makes two archival copies in well-established, standard magnetic formats; the copies are then stored in two separate locations, one of which is environmentally perfect. Only after archival copies are made is there further processing, which includes, for example, examination of the dataset to determine if documentation is complete. Once processing is done, there is a new version of the dataset, called the servicing version, and this is archived, as well, on magnetic tape. The version available to the public is on magnetic disc.

ICPSR thus maintains two separate collections of data, one for servicing clients and one for ensuring the archival integrity of the collection. Each collection is managed by an independent database,

and the archiving collection and its associated database are accessible only to Archival Operation staff. The archival collection is routinely migrated from medium to medium but is used only when a dataset in the servicing collection fails. Moreover, an incremental backup is done every evening, and a full backup once a week. With all of these fail-safe mechanisms in place, ICPSR can boast that it has never lost a dataset.

Many people in libraries know about ICPSR's archiving and servicing of datasets related to faculty research studies, but more of their revenue is actually derived from archiving and servicing for contract clients, many of whom are federal government departments. An example is the crime and justice archives of the Department of Justice.

Archival Operation costs nearly a quarter of a million dollars to run every year, with a like amount paid every time a major migration of data takes place. These costs do not include organizational overhead. The unit processes some 5,000 datasets a year. Rockwell suggests that they have been able to keep costs down for two key reasons: 1) by focusing on one basic product and service, they are able to get better and better at what they do; and 2) they take every possible advantage of the increasing density of storage media.

Though quite focused in scope, ICPSR is a fascinating operation for libraries looking at the opportunities and challenges of digital archiving; in a setting where most others are still planning, they are doing it.

Contact:
Richard Rockwell
Executive Director
ICPSR
313-764-2570
<richard@icpsr.umich.edu>
<http://www.icpsr.umich.edu>

OCLC Focuses on Members' Current Needs in Digital Preservation

As a membership organization, OCLC is currently responding to members' immediate needs in the archiving of digital information, deferring for the moment intensive work on the preservation of digital-only information. At the same time, several projects will function as learning laboratories as the cooperative utility moves into the digital-only arena.

OCLC member needs are reflected in two well-documented OCLC initiatives that focus on both older special collections and on current journals and their back files. In the former, OCLC has invested approximately $500,000 in electronic archiving projects to test

models, technologies, and capabilities. For example, several collections, such as selected materials from the New York Public Library's Schomburg Collection, are being scanned and stored digitally. In designing this project, OCLC set out to learn how to capture large collections and make them accessible in economical ways. They have deliberately chosen a broad range of materials, including monographs, photographs, and newspapers, which will be made available through the OCLC FirstSearch system.

The second project, OCLC FirstSearch Electronic Collections Journals Online, will ultimately make thousands of journals available online, supplying both subscription and document delivery access. The program also has committed to perpetual preservation storage of all journal back files; that is, subscribers will have perpetual access to the backfiles they originally subscribed to.

"It is our everyday business to archive electronic resources, migrate files—everything that goes into preserving digital information," says Donald J. Muccino, OCLC Executive Vice President and CEO, referring to the ongoing maintenance of their vast bibliographic databases. "In many ways, OCLC is uniquely positioned to help libraries deal with electronic archiving. We have over 26 years experience in migrating WorldCat (the OCLC Online Union Catalog) through constantly changing storage technology. Now we are building the technical infrastructure for electronic archiving and exploring the economics of preserving digital content for direct use by our members." That technological infrastructure includes eight "silos" in OCLC's computer room that use robots to store, retrieve, and mount tapes. This "near-line" storage is faster than offline tape and less expensive than online disk and allows OCLC to mount large files of full-text and image information for on-demand retrieval.

As with any cost-recovery operation, OCLC must be constantly mindful of the economics of new ventures. With all of their digital archiving efforts, including digital-only, how will they recoup costs? Will costs be embedded in membership fees? Will individual members subscribe to sets of titles? Of particular interest are the patience levels of end users. It can be much less costly, both for OCLC and the member, to store digital information in near-line storage devices "offline," but such storage means waiting several minutes—perhaps even hours—for delivery to a desktop. To minimize the wait time, a mix of storage devices will be employed to allow the management and optimization of where, and therefore, how quickly items can be retrieved, based on age, type, and last use of the material. Will such a wait be considered a boon (as compared with the weeks and months that such transactions used to take) or will it be considered unacceptable (as compared with the virtually instant access that systems are currently capable of delivering)? Moreover, if something is

asked for, what are the chances that it will be asked for again very soon? The answers to questions like these will shape the architecture of digital storage at OCLC in the next few years. OCLC will be answering many questions about archiving special collections and journals (costs, response time, storage array, etc.) during the pilot project by working with contributing content participants and libraries/users around the world.

According to Muccino, OCLC will be only one site member in an organized partnership of distributed digital collections and services to be developed within the next few years. Some participants will be agencies such as OCLC, some will be scholarly associations, some will be libraries. This is why he feels that the work of organizations such as the Digital Library Federation and the Coalition for Networked Information is so critical: his hope is that they will coordinate a sorely needed national effort.

Contact:
Donald J. Muccino
Executive Vice-President and Chief Operating Officer
OCLC
614-764-6034
<djm@OCLC.org>

RLG Experiments, Develops Infrastructure

As co-sponsor with the Commission on Preservation and Access of the seminal report, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* (May 1996), available at <http://www.rlg.org/ArchTF/index.html>, the Research Libraries Group has mounted several initiatives to bring the report's key recommendations to reality. Two of these efforts are well documented on the World Wide Web: Studies in Scarlet and the Global Immigration Project. Each seeks to take a different look at the feasibility and economics of centralized and distributed models for digitizing archival materials. Both also achieve an important criterion: they meet member needs by providing an opportunity for collaboratively gaining digital reformatting experience, creating a body of recommended practice for future initiatives, and developing important research collections for a broad audience of scholars.

Studies in Scarlet grew out of the need for content to serve as a test-bed for RLG's developing archive and access infrastructure, ARCHES. Focusing on marriage and the law, the project gathers largely published materials from 1815 to 1914 from the U.S. and the U.K. The organizational model for this project is centralized (digitized materials are submitted by seven participating institutions and

stored at RLG), selective (rather than combining whole collections from the contributing institutions), and primarily focused on access.

The Global Immigration Project grew out of a need expressed by member collections officers to preserve and make available collections pertaining to immigration that are difficult to access or are slowly deteriorating. This project is organized in a different way: it is distributed (digitized collections are stored at 11 participating institutions in the U.S., the U.K., and Canada), comprehensive (whole collections and/or major portions of collections are being digitized), and primarily focused on preservation quality capture.

Though basically page image-based projects involving largely public domain text materials, these are two very important projects for the library community to watch during the next few years.

As they face the problem of archiving information that exists in digital form only, RLG is focusing on building ARCHES, a truly functional archival server infrastructure. James Michalko and Ricky Erway of RLG point out that the problems in this area are less technical than legal and financial. Most digital-only materials are copyrighted and cannot be archived without permission. Beyond copyright is the very real question of the information owners' willingness to pay for long-term preservation storage—or the users' willingness to pay for access to those materials. While RLG is a not-for-profit organization, most of its services need to be self-supporting.

With the ARCHES architecture, RLG is able to provide: access via the Web, storage of digital objects, a system for providing uniform resource names, a document authentication system, a license server, and full-text searching capability—all of which will be essential as they move toward providing archiving and access for digital data. A significant piece still under development is the actual storage and maintenance software. Says Erway, "When the Studies in Scarlet Project is finished, we will have about twenty-five gigabytes of data on the archival server, enough for us to learn some very important lessons about providing digital storage services for members and other information providers."

Contact:
James Michalko
President
RLG
650-691-2240
<bl.jpm@rlg.org>

Ricky Erway
Member Services Officer, Digital Initiatives
RLG
650-691-2228
<bl.rle@rlg.org>

National Libraries and Archives

LC Coordinates Management of Electronic Resources

Most librarians are familiar with the Library of Congress National Digital Library Program (NLDP), which is in the process of converting some five million images into digital form. Not so familiar, perhaps, is a significant enterprise-wide effort, called the Electronic Resources Project, aimed at coordinating LC's approach to electronic resources in its many organizational units and operations, including the NDLP, the Copyright Office, Congressional Research Service, Library Services, Law Library, and Information Technology Service.

Like libraries many times smaller, the Library of Congress is also building a digital infrastructure and is looking to bring consistency to library-wide practice. At the present time, the library's Electronic Resources Project is focusing primarily on the acquisition, collection development, and use (access) of electronic resources. In the process of guideline development, the Library of Congress is creating a template that may prove very useful to others who embark on such planning.

Several options are under consideration, among them expansion of the existing collection policy guidelines to incorporate online and other formats of electronic resources, with special focus on defining "best editions" and preferred physical and electronic formats for permanent retention. The acquisitions component of the developing guidelines includes intellectual property rights management, identification of problems unique to acquiring electronic resources, and identification of selection tools and techniques.

Of special note for those interested in digital preservation is LC's consideration of several possibilities: working with national and international groups to establish standards for archival repositories for electronic resources; establishing internal standards for preservation copies (including policies on when to refresh, reformat, or migrate, and how to treat extensions of digital resources such as hyperlinks); determining LC's responsibility as a library of last resort; and deciding how to treat dynamic works.

The Electronic Resources Project Study Group is examining and will make recommendations on a number of issues, including how LC should best organize and staff itself to deal with these new forms of library collections. As an enterprise that feels acutely its responsibility as a national library, the library is being especially careful in its consideration of what for other libraries might be the standard options for digital and electronic preservation and access, including the option that LC itself might not own and manage the preservation of everything it needs to have access to.

While the Library works on electronic policy issues, yet another project moves forward in its preservation directorate—a project that illustrates the alarming problems of the digital future only too vividly. Sound recordings made to be played on equipment that is already obsolete are being transferred into digital form in a new state-of-the-art audio facility. Diane Kresh, Director for Preservation, characterizes this effort as "scrambling to stay ahead of obsolescence." Experience has taught Kresh and others at LC that early hopes of scanning everything and making it accessible must be tempered; a more careful strategy calls for scanning once for both preservation and access purposes.

LC hopes to have the Electronic Resources Project completed early in 1998, at which time copies of resulting documentation will certainly be of interest to the rest of the library community. A brief account of LC's Legal Deposit Program and its approaches to acquiring digital materials is included in the next section.

Contact:
Diane Kresh
Director for Preservation
Library of Congress
202-707-6072
<dkresh@loc.gov>

National Library of Canada Focuses on Policy, Acquisition Procedures

The National Library of Canada (NLC) accepts electronic publications in its depository program but is still working out a number of policy and procedural issues. Of special interest among the documentation emerging from the NLC are collection development priorities for electronic materials.

Of course, NLC's primary focus in collecting and archiving is on Canadiana, including government publications. Within those parameters, the Library's current policy is to concentrate first on current electronic publications not available in any non-electronic format; second on complete digitized "reprints" of non-electronic materials; and third on complete electronic publications that parallel current publications in other non-electronic formats. Though they have established further priorities (e.g., works in progress containing sufficient material to stand on their own), they are not yet acquiring beyond the first three criteria. Further, NLC prefers "structured publications in standard, non-proprietary formats," and they are not collecting OPACs, databases, A&I Services, public communications sent over networks (e.g., e-mail, listservs), and World Wide Web or Gopher sites.

Though many materials are routinely deposited in electronic form, finding appropriate materials is also the job of library subject specialists,

who regularly scan the Internet for such materials. After an extensive pilot project to test the implications of acquisition, access and preservation of electronic publications, NLC has moved systematically to acquire government and non-commercial electronic publications. Further examination of the implications of legal deposit is underway.

As they work on developing and building their electronic library, NLC has identified several successful strategies. One is to acquire materials in standard formats, such as SGML, as much as possible. Another is to put as much of their discourse as they can into the print world analogy—a mode which is more comfortable than a digital one for many people. Electronic collections are accessible at the NLC web site and provide full-text search capability.

Contact:
Nancy Brodie
Government Information Holdings Officer,
Information Resource Management,
National Library of Canada
613-947-5887
<nancy.brodie@nlc-bnc.ca>

CER/NARA Mission: To Provide Ready Access to Essential Digital Evidence

The Center for Electronic Records of the U.S. National Archives and Records Administration (NARA) is yet another operation with substantial experience with archiving digital information. Under NARA's umbrella mission, the Center currently acquires some 10,000 files a year, up from 159 nearly a decade ago. Acquisitions are doubling every two years, a rate that will move them to handling a million files annually in six or seven years.

The Center has chosen magnetic tape as its archival medium, even though it is not a preservation medium: its advantages—being so well known and commonly used in the field—far outweigh its disadvantages. Magnetic tape also enables the archival manager to separate error detection and error correction. Other media do not provide such separation, so that crashes occur before errors are detected.

Though physical preservation is important, a much more difficult challenge, according to Director Kenneth Thibodeau, occurs at the higher conceptual and logical levels, and there is a constant process of educating agencies who submit data. Fully one-half of submissions, principally databases and natural language documents at this point, come with problems that have to be worked out: for example, documentation may be lacking or there may be a discrepancy between the description of data content and the actual content itself. The good news is that, as agencies become better educated, problem submissions decrease.

One reason problems occur is that agencies vary greatly in the time it takes them to submit data after its creation to the Center, with some transmitting immediately and others holding transmission off "forever." A key problem in prolonging the gap between data creation and submission is getting data out of the older systems in which they were born. The Center, of course, urges prompt submission as the best strategy, but it also assists agencies, when possible, with data extraction.

With the number of submissions quickly growing each year, the Center is happy to report satisfying increases in productivity through strategies such as routinizing within classes of records. For example, they are looking to manage e-mail, for which they are currently doing test transfers only, as a special type of database. Another basic strategy is the decision not to provide processing capabilities to researchers along with the data they request; the assumption is that users will have these means at their disposal.

Thibodeau stresses the public service mission of the Center, which provides access that is as open and liberal as possible to the archived electronic record of the federal government.

Contact:

Kenneth Thibodeau
Director, Center for Electronic Records
National Archives and Records Administration
301-713-6635
<ken.thibodeau@arch2.nara.gov>
<http:www.nara.gov>

Legal Deposit Arrangements

The Conference of Directors of National Libraries (CDNL) has taken an intensive look at the legal deposit of electronic publications. A 1996 report of a CDNL Working Group chaired by Dr. Brian Lang, Chief Executive of The British Library, states: "One of the main responsibilities of national libraries is to assemble comprehensive collections of the published outputs of their nations, record and organize them for use, and preserve them for posterity. National libraries have been able to develop their print collections through legal deposit, a statutory provision which places a legal requirement on producers of publications to deposit their works in designated institutions... [t]he CDNL believes strongly that governments should provide statutory rights to enable national libraries to receive non-print publications through legal deposit." The CDNL Working Group report contains guidelines to assist national libraries to make their cases for the legal deposit of electronic publications—guidance with potential applicability to other kinds of libraries and archives.

The CDNL report (*Conference of Directors of National Libraries: The*

Legal Deposit of Electronic Publications. UNESCO, 1996) also summarizes the status of such depository programs in a number of nations. Most depository programs are still wrestling with the initial challenges of archiving electronic information—developing definitions, policies, and priorities. A few of these efforts are summarized here.

Australia is particularly active in this area, with key agencies in 1995 proposing that legal deposit provisions be extended to include formats previously not included, such as electronic resources. In anticipation of changed regulations, the National Library "has been receiving some electronic publications on CD-ROM and floppy disc under a voluntary deposit scheme" with publishers. During 1996, the Library began pilot projects to preserve Australian online publications. At its Web site, *Managing Electronic Records - a shared responsibility* <http://www.aa.gov.au/AA_WWW/AA_Issues/ManagingER.html>, the Australian Archives provides a continually updated report on its efforts to organize management of national electronic records. At present, the Archives prefers government agencies to be responsible for their own archiving, since "individual agencies are most likely to understand their electronic systems and the specific applications required to maintain the records they contain." The Archives sees itself as assisting agencies in several ways, however: in the identification of records of lasting value; in the planning for disposal of records of no lasting value; in the identification of contextual and metadata information that must be retained for effective access; and in the development of access- and record-keeping systems.

Canada is working on a similar extension of depository provisions. The National Library is currently identifying, through the Internet, "Canadian online journals, texts and databases and keeping copies, as appropriate, in either printed or electronic form" as part of a cooperative project with publishers. Changes in deposit regulations that would accommodate electronic resources await the resolution of key legal and copyright issues.

In **Denmark**, "electronic datasets produced for public administration are deposited in the National Archives and electronic research results are held by the Danish Data Archive."

Norway's depository regulations are only five years old, but already it has organized to accept deposits of both static and dynamic electronic documents and is managing storage on gold-plated CD-ROM, to which deposited documents are migrated.

The **United Kingdom** is in the process of expanding deposit regulations to include electronic resources. A proposal from the British Library in 1996 recommended that new legislation should be comprehensive but that "implementation should be incremental, starting with well-established publishing media, and generally taking on board the newer media publications."

In the **United States**, copyright legislation differs from most others in one key respect: "since copyright exists from the moment of creation of a work, all published works must be deposited, whether registered or not." Mandatory deposit regulation "of some categories of non-print materials, including machine-readable works and CD-ROMs but not including online databases are contained in the 1993 *Code of Federal Regulations*." Presently, the U.S. Copyright Office "considers that all types of material are subject to mandatory deposit, but lacks legal authority for collecting certain materials" because the concepts of "publication" and "transmission" are not closely defined for online publications.

The major issues faced by libraries in pursuing the deposit of electronic publications were the focus of a study undertaken for the European Commission by J. S. Mackenzie Owen and J. v. d. Walle of NBBI, Project Bureau for Information Management in the Netherlands. The conclusions and recommendations of the study were discussed by librarians and publishers at a workshop held in Luxembourg in December 1995, where it was agreed to establish a task force to work together to develop collaborative agreements and practical test beds. A substantive report describing the approach being taken by European agencies to archive digital publications was published by the European Commission (*Depository Collections of Electronic Publications*, by J.S. Mackenzie Owen and J. v. d. Walle. European Commission, DG XIII-E/4. Luxembourg, 1996. ISBN 92-827-7643-3).

Publishers

The American Institute of Physics Plans for Perpetual Preservation

The American Institute of Physics (AIP), an umbrella organization of not-for-profit physics societies, is an example of a scholarly association that is making the transition to an electronic future in decisive ways. The Institute publishes eight "archival" or research journals, eight Russian translation journals, and three non-archival periodicals, as well as conference proceedings that are expected to number between 35 and 40 publications in 1998. Recently, they have been working on plans to archive the journals, which have been available online for about a year. Archival plans call for multiple sites, regular refreshing and migration, and procedures for caring for the archives of ceased publications. For the present, the paper edition will continue to be the edition of record.

These electronic archives will be available to all continuing subscribers, as long as they maintain active subscriptions, and to others on a cost per transaction basis. The archives will provide links—both

backward and forward—to abstracts of other publications of AIP and its ten member societies held in the searchable database known as SPIN. Lapsed subscribers will have access to physical copies of back files for the years they subscribed, in either paper or CD-ROM format, at cost. These copies will not, obviously, have the full functionality of electronically accessible files.

For more than twenty years, the AIP has offered its Physics Auxiliary Publications Service, which is now also available in electronic form. Through this service, subscribers can access data related to published research that has been deposited by authors with the AIP.

In addition to the links provided to other AIP and member-society publications, the Institute also provides links to INSPEC citations. Before providing such links, editors test all of them, as they have found a number of errors in citations provided by authors; when errors will prevent access, the links are corrected.

In response to the doubts expressed by some librarians about publishers' commitments to preserve their electronic publications forever, John Scott, AIP's Journal Publisher, responds that professional societies will have a better chance of surviving—and continuing to preserve digital archives—than commercial publishers.

Contact:
John Scott
Journal Publisher
American Institute of Physics
516-576-2297
<jscott@aip.org>

UMI Dissertations is Recreating Its Business for the Digital Age

Through its new ProQuest Digital Dissertations Initiative, UMI, formerly University Microfilms International, is testing new ways of conducting its bread-and-butter business: providing access to the dissertation literature. For most dissertations, UMI starts with microform copies, its traditional mode of distribution. Microforms are then scanned and translated into PDF files, which can then be made available to purchasers on the World Wide Web.

This new mode of service provision is starting in 1997. Because of relatively low demand, backfiles of dissertations will remain in microform, though selected items from the backfiles that are in significant demand may eventually be migrated to Web-accessible format, as well.

UMI accepts basically text-based dissertations from authors in digital formats, including Word, WordPerfect, Adobe PDF, TeX, and LaTeX 97, the latter two being particularly useful in the sciences for recreation of formulae. When dissertations are in "compound form," both text

and other media, they negotiate with the authors to provide digital materials, e.g., musical clips, on storable media, such as CD-ROMs.

For long-term preservation, UMI will continue to depend on preservation microfilming as an alternative format. Digital submissions will be stored on both tape and CD-ROMs. These digital preservation copies will be sampled on a regular basis and refreshed as read errors are found. Parallel preservation copies are located in different vaults for security purposes.

Because they are now shifting to a largely digital service mode, UMI is repositioning its understandings with graduate schools and reworking the templates they use with authors. New technology enables UMI to offer a group of enhanced services to graduate institutions and libraries. One such service provides free Web access to anyone at a specific institution to any of that institution's dissertations simply by using the appropriate IP address. Another capability is their "Current research @. . ." service; using this search mode enables the inquirer to find out about the research that has been done recently at a specific institution.

Yet another service in development is the posting of announcements of dissertations defenses that will be provided by the graduate schools. It is hoped that such a service will increase interest in dissertations and the defense process—and may even be used by prospective employers in seeking talent. Libraries are likely to find these services useful in their liaison and collection development programs.

Finally, libraries should benefit significantly from UMI plans to supply submitting institutions with MARC cataloging records for the documents they publish through UMI. All of these services will be provided to the individual authors and graduate institutions at no additional cost.

Contact:

William Savage
Director, Academic Relations
UMI Dissertations Publishing
313-761-4700
<bsavage@umi.com>

Coordinative Efforts

Peter Graham of Rutgers Plans ARL/CNI/CLIR Workshops, Initiates Local Discussions

Though most Reports from the Field focus on institutional efforts, occasionally the work of a single individual merits attention. Peter Graham, Associate University Librarian at Rutgers, has recently

submitted a plan to the National Endowment for the Humanities for ARL, CNI, and CLIR to sponsor a series of workshops on the topic of digital archiving. The concept behind the five workshops, which, if funded, will be offered starting in June 1998, is that each will "devote itself to a particular topic with the aim of noting what progress has been made so far, specifying what the major problems in the topic are and what would be fruitful means of investigating the problems and solving them." Though content is not set in stone, topics proposed so far include: 1) archival contents and redundancy (collection development issues), 2) technological obsolescence and migration techniques, 3) current and future hardware technologies, 4) authenticity (integrity, intellectual preservation), and 5) intellectual property. Advisors to this project include Duane Webster, Clifford Lynch, and Deanna Marcum.

Graham has also drafted a discussion document aimed at opening dialogue within his own Rutgers University Libraries (RUL) on the subject of digital archiving strategies. In his paper, Graham first describes the local climate for digital preservation, a mix of recent achievements and opportunities. He then moves to two propositions: 1) that the libraries should consider development of "interdependent policies and practices for the archiving of networked electronic information," and 2) that processes be developed to initiate "RUL's role in digital archiving." Specific suggestions are that a working policy statement be developed and a work group be named as ways to get started.

Graham's continuing interest in the issues related to electronic archiving have also prompted him to maintain a very useful bibliography on the subject at <http://aultnis.rutgers.edu/texts/ElectLibBib.html>. Additional information about the workshops proposal is available at <http://aultnis.rutgers.edu/DPAW/dpawhome.html>.

Contact:
Peter Graham
Associate University Librarian for Technical and Networked Information Services
Rutgers University Libraries
732-445-5908
<psgraham@rci.rutgers.edu>

The Digital Library Federation Moves to Coordinate National Effort

The Digital Library Federation (DLF), formerly the National Digital Library Federation, under the organizational umbrella of the Council on Library and Information Resources and the leadership of a new Director, Donald Waters, seeks to move forward the agenda

identified by the Commission on Preservation and Access (CPA) and the Research Libraries Group (RLG) in their landmark report, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* (1996). Constituted a year ago as a charter organization with twelve (now fourteen) university library members, the National Archives and Records Administration, the Library of Congress, and the New York Public Library, DLF is currently moving to expand its membership and its coordinating influence.

As noted in the *CPA Newsletter*, a DLF Task Force last year identified three areas of inquiry:

- Discovery and retrieval: "the heterogeneity of the information available in digital form—different data structures, search engines, vocabularies for access—significantly challenges users in their ability to identify and retrieve needed resources."
- Intellectual property rights and economic models: "most of the technical requirements for the management of intellectual property rights are now, or will shortly become, available."
- Archiving digital information: "the greatest test of adherence to the goal of creating a national digital library very likely is a commitment to preserve culturally significant digital information as part of the national heritage."

Within these three areas, the DLF Task Force recommended: 1) lowering the barriers to access for digital materials and providing cross-collection search capability; 2) putting in place a clear and articulate policy to regulate rights relationships among federation institutions; and 3) fostering and facilitating a commitment to digital archiving among participants.

DLF has a key role in: 1) helping to focus technology research and development; 2) promoting dialogue among libraries, archives, museums, government agencies, and major technology vendors to determine standards and best practices; 3) fostering demonstration and prototype projects to define and solve problems before large-scale implementation; and 4) enabling different kinds of institutions to create digital library information resources that are genuinely useful for education and research because they are "consistent, coherent, and of high quality."

Contact:
Donald Waters
Director, Digital Library Federation
Council on Library and Information Resources
203-498-6076
dwaters@clir.org>

Additional Resources: Preserving Digital Information

Printed and Web site resources are intermixed in the following list. The first two citations are excellent means of getting started on the subject of preserving digital information.

Bibliography on Electronic Library/Digital Library Issues. This Web site is maintained by Peter Graham, Rutgers University Library; Section D contains citations on preservation and archiving issues. Link from here to a variety of Web sites. Available at <http://aultnis.rutgers.edu/texts/ElectLibBib.html>.

Commission on Preservation and Access Newsletter. This Web-based newsletter provides coverage of digital preservation issues, among other subjects. Available at <http://clir.org/cpa/newsletter>.

Electronic Records Research and Development: Final Report of the 1996 Conference held at the University of Michigan, Ann Arbor, June 28-29, 1996. The *Report* is available online at <http://www.si.umich.edu/e-recs/Report>.

Gardner, James B. *Report on Documenting the Digital Age.* History Associates Incorporated, 1997.

Managing Electronic Records – a shared responsibility. This Web site provides some insight into how digital preservation policy is being developed in Australia. Available at <http://www.aa.gov.au/AA_WWW/AA_Issues/ManagingER.html>.

Marcum, Deanna B. "The Preservation of Digital Information." *The Journal of Academic Librarianship* 22 (Nov. 1996): 451-54.

Michelson, Avra, and Jeff Rothenberg. "Scholarly Communication and Information Technology: Exploring the Impact of Changes in the Research Process on Archives." *American Archivist* 55 (Spring 1992): 236-315. Reprinted as *Rand Reprint* number 187, Rand Corporation.

National Media Laboratory. Web site, available at <http://www.nml.org>. Of interest is *Long-Term Preservation of Digital Materials*, prepared by John W. C. Van Bogart. Presented at the National Preservation Office (NPO) Conference on "Preservation

and Digitisation: Principles, Practice and Policies" (September 3-5, 1996, at the University of York), this talk is specific to librarians and archivists who need to deal with magnetic tape and optical disc as an archival medium. Available at <http://www.nml.org/Publications/Presentations/PreservationOfDigitalMaterials/>. Also of interest is *Magnetic Tape Storage*, a talk presented by Van Bogart at NARA's 11th Annual Preservation Conference, March 14, 1996. Available at <http://www.nml.org/Publications/Presentations/MagneticTapeStorage/>.

National Research Council. *Preserving Scientific Data on Our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources*. Washington, D. C.: National Academy Press, 1995.

Owen, J. S. Mackenzie and J.v.d. Walle. *Deposit Collections of Electronic Publications*. Luxembourg: European Commission (DG XIII-E/4), 1996.

PADI: Preserving Access to Digital Information. This Web site provides excellent coverage of developments in Australia related to digital preservation. Visit <http://www.nla.gov.au/dnc/tf2001/padi/digital.html>.

Preserving Digital Information: Report of the Task Force on Archiving of Digital Information. Commissioned by the Commission on Preservation and Access and the Research Libraries Group. Washington, D.C.: CPA, 1996. A very substantive, detailed introduction to the subject and the issues. Available on the Web at <http://www.rlg.org.ArchTF/>.

Rothenberg, Jeff. "Ensuring the Longevity of Digital Documents." *Scientific American* 272.1 (Jan. 1995): 42-7.

Sanders, Terry. *Into the Future: On the Preservation of Knowledge in the Electronic Age*. Produced by the American Film Foundation, the Commission on Preservation and Access, and the American Council of Learned Societies. "This program confronts the hidden crisis of the digital information age—will the human record, as it is increasingly stored in fragile and complex digital forms, survive into the future?" (PBS program notes, <http://www.pbs.org/whatson/1998/01/descriptions/INFU.html>). A review of the film by Paul Wallich is available in the January 1998 issue of *Scientific American*, <http:www.sciam.com/1998/0198issue/0198review2.html>. The film can be ordered from the American Film Foundation, contact: 310-459-2116.

Society of American Archivists (SAA). *Statement on the Preservation of Digitized Reproductions*. Available at the SAA Web site: <http://www.archivists.org/governance/resolutions/digitize.html>.

The Scientific Electronic Archive. Web site. Summary of a discussion session held at the ICSTI (International Council for Scientific and Technical Information) 1997 General Assembly, June 6-7, 1997, Philadelphia, PA. Available at <http://www.icsti.nrc.ca/icsta/archive.html>. This ICSTI discussion session on the scientific electronic archive was a follow-up to the ICSU Press–UNESCO Expert Conference on Electronic Publishing in Science (February 19-23, 1996); see <http://www.lmcp.jussieu.fr/icsu/Information/Meetings/index.html>.

U. S. National Archives and Records Administration (NARA). *Ready Access to Essential Evidence: The Strategic Plan of the National Archives and Records Administration*. Web site. "Impressive developments in technology for creating records have not been matched by technological developments for matching them." Available at <http://www.nara.gov/nara/vision/naraplan.html>.

———. *Title List: A Preliminary and Partial Listing of Data Files in NARA*. The *Title List* has entries for approximately 13,000 of the more than 34,000 electronic records files in the custody of NARA (<http://www.nara.gov>). Available at <http://gopher.nara.gov:70/1/inform/dc/electr/titlelst/>.

University of Michigan. *D-Lib Magazine*. Several relevant articles can be found in this e-publication, which can be found at <http:www.dlib.org>. See especially: <http:www.dlib.org/dlib/may97/michigan/05pricewilkin.html>; <http:www.dlib.org/dlib/july97/humanities/07/powell.html>; and <http:www.dlib.org/dlib/july97/america/07/powell.html>.

SYSTEMS AND PROCEDURES EXCHANGE CENTER

Flyer 228

Preserving Digital Information

Transforming Libraries 5: December 1997

Introduction

THE PROBLEM of preserving digital information and the strategies that are and might be employed to address it are the focus of this fifth issue of *Transforming Libraries*. In laying out the problem, a number of questions are asked: How will we transmit research results, official evidence, and our cultural heritage to future generations? Supposing we copy new data onto new media, will we have the software needed to make sense of the data? Who will pay for preserving digital information? Who will decide what to preserve, on what media, and with what level of metadata management? Answers to all of these questions will not likely be available for some time although the technical issues may be solved more readily than some of the policy issues.

Twenty-one individuals involved at the technical or policy level in developing strategies for preserving digitial information were interviewed for this publication. As reported in the "Issues Roundtable" section there is consensus on a number of points including:

- The problem is growing faster than we are currently working to solve it.
- Often confusing is the difference between digitizing for preservation and digitizing for access, and preserving digital-only information.
- More analysis of the problem, including further assessment of the technical issues, costs, options, and risks needs to be undertaken in a collaborative manner.

Key Models

Currently, as the major players are working through the choices for policy development, standards setting, and the economics of preserving digital information, a number of models are emerging. Ten stakeholder groups who will play a role in designing and testing possible models are: researchers or originators of information; publishers; repositories, libraries and museums; the national libraries and archives who operate legal deposit programs; the cooperative service agencies, e.g., OCLC and RLG; private storage providers; the scholarly associations; indexing and abstracting services; segmented market providers, e.g., UMI and ICPSR; and certified digital archives as proposed in a study by the Commission on Preservation and Access and the Research Libraries Group.

Reports from the Field

A regular feature in the *Transforming Libraries* series is the "Reports from the Field," which provides a snapshot view of how a number of organizations are approaching the problem at hand. While just about all the players agree that the most pressing problem is how to preserve the digital-only information, i.e., information that has

only existed in electronic form, this report features a mix of activities or experiments in preserving digital-only information as well as conversion to digital form for preservation purposes.

Featured are some of the varying institutional approaches to preserving digital information:

- The Cornell University and the National Agricultural Library Partnership, the National Preservation Program for Agricultural Literature.
- Harvard College Library's technical and tactical approach to digital library development.
- Just-in-case and just-in-time planning for digital preservation and access at the University of Michigan.
- A long history of regular, periodic migration at the Inter-university Consortium for Political and Social Research (ICPSR).
- Responding to members needs is the focus of current operations and planning for the future at OCLC, while at RLG the focus is on building an infrastructure through projects and experimentation.
- In addition to the well publicized National Digital Library Program, the Library of Congress is involved in a number of initiatives, such as the Electronic Resources Project that focuses primarily on acquisition, collection, development, and use of electronic records, which includes working with national and international entities to address standards for electronic archival repositories.
- While working to build the electronic collections at the National Library of Canada, staff are currently focusing on acquiring government and non-commercial materials, and they also are aiming to acquire materials in standard formats such as SGML.
- Digital acquisitions at the Center for Electronic Records of the U.S. National Archives and Records Administration (NARA) doubles every two years as more and more agencies are submitting electronic records.
- The Digital Library Federation, a program of the Council on Library and Information Resources, is moving forward the recommendations made in the CPA/RLG report, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* (1996).

Additional Resources

This issue of *Transforming Libraries* includes an extensive list of additional resources–mostly accessible via the WWW. They include articles and publications as well as other projects or organizations who are also addressing the issue of preserving digital information.

This SPEC Kit was written by George Soete, Transforming Libraries Editor, and Mary Case, Director, ARL. Office of Scholarly Communication.

The electronic component of this issue can be found at <http://www.arl.org/transform/pdi/>.

SPEC Flyer (ISSN 0160 3574) **©1997 by the ASSOCIATION OF RESEARCH LIBRARIES, OFFICE OF MANAGEMENT SERVICES. ARL grants blanket permission to reproduce this information for educational use, provided that the source, author, and copyright notice are included. Send requests for commercial use to the ARL Publications Department, 21 Dupont Circle, Washington, DC 20036. (202)-296-2296. SPEC Kits and Flyers are available by subscription and single issue.**

innovations issues
Association of Research Libraries
Office of Management Services
21 Dupont Circle
Washington, DC 20036
43